

Dowding, H. (2016). Open Access, Digital Preservation, and the Changing Scholarly Record. In Smith, K. & K.A. Dickson (Eds.), *Open Access and the Future of Academic Libraries*. Lanham, MD: Rowman & Littlefield.

### **Open Access, Digital Preservation, and the Changing Scholarly Record**

*The discovery of future knowledge is a common good and a treasure we owe to future generations. The challenge of today's generation is to keep the pathways to discovery open.*

-Hess and Ostrom

In a prescient 1990 article in *American Archivist*, geography professor Kenneth Foote discusses the role of archives in the building of a collective cultural memory, using as an example the work of the Human Interference Task Force. Convened in 1980 by the U.S. Department of Energy, the group was tasked with safeguarding nuclear waste deposits, a role that included communicating the location, management, and threat of nuclear waste to future generations. With radiation's half-life spanning ten millennia, the threat of radioactive poisoning must be conveyed across so many generations that even things like shared language can no longer be assumed. The task force had to contend with conveying crucial information to a population about which they knew nothing.

The challenge of long-term nuclear waste management illustrates the fact that we as curators of the scholarly record must anticipate not only the needs of those within our institutions at this moment in time, or even the needs of their children, but also the needs of those ten thousand years from now. They will not be born for centuries, they will have technologies that we cannot even fathom, and the ways in which they choose to use the artifacts of our current culture is yet to be determined. The only thing we know about them is what we know about ourselves: we build our knowledge of the past based on those items that have survived, either by accident or by intervention, to wind up in our hands. It is with this background in mind that I will begin to build the connections between open access and digital preservation, and their relationship to the evolving scholarly record.

Conceptually, access and preservation are often viewed as dichotomous; preservation involves removing things from public view, whereas access involves a certain amount of risk in putting things out where they can potentially be harmed. In the case of digital scholarship, however,

access and preservation go hand in hand, and are often paired with the concept of curation as a bridging factor (Borgman, 68). While access is a broad term, encompassing technical, social, and legal issues, and while it is by no means exclusively associated with the open access movement, it is important to begin a discussion of the linkages between open access and digital preservation by recognizing the fact that they are already intrinsically connected within this trinity of issues. Without preservation, there can be no access; without access, preservation is impossible. Many scholars have drawn links between the concepts of digital preservation and open access, whether overtly or not. For example, as the main facets of the “infrastructure of openness,” Tatum and Jankowski cite “standards, content interoperability, and levels of customization” (185), all areas of deep concern to the digital preservation community. The threats to what Hess and Ostrom refer to as “knowledge commons” are quite similar, and include “pollution and degradation, and nonsustainability” (5). Corrado and Moulaison draw a direct connection between open access and digital preservation, stating that their shared interests encompass “openness, transparency, interoperability, quality, and sustainability” (188). Whatever the micro aims of digital preservation and open access, the macro result of both is a more comprehensive scholarly record.

Since many scholars see open access as inevitable rather than merely possible (OAMHSS, 7), it is necessary at this point to begin a discussion of its impact on the enduring scholarly record. While open access presents new pathways for inclusivity and allows for greater access to scholarship, the concern of current business models rests almost entirely on the costs that are incurred at the beginning of a digital scholarly object’s lifecycle. Should this lack of consideration for the ongoing management of scholarship in open access models be cause for concern? Or is it simply a relic of older publishing models? In the case of both open access and digital preservation, as in nearly every conversation held in the academic library, the discussion of long-term viability must begin with cost.

### **Economics of the Open Scholarly Record**

Current research into the cost and value associated with an open access scholarly record is divided into studies of open access, on the one hand, and studies of digital preservation, on the other; there are not many studies exploring the overlap. In terms of open access, while there is no shortage of studies related to economics, the research is still relatively in flux. This is largely

because the concept of value in scholarship is indirect and hard to define, as those commodifying scholarship are rarely the same entities funding its initial creation. Until now, value was understood mainly in terms of prestige and promotion and tenure; however, these bases are shifting as academia evolves. Willinsky notes that the value of scholarship is in its dissemination (2009, 53), which can arguably serve traditional purposes but also provides other benefits such as preservation and collaboration. Burns, Lana and Budd (2013) approach the value of open access scholarship from the library's perspective, citing as a major benefit its ability to influence scholarly communication by centralizing management.

On the cost end, because open access is undertaken as part of daily functions, it is hard to quantify the costs involved in creating one scholarly object; however, several reports provide a granular understanding of budgets. For example, Houghton's 2009 study of alternative scholarly publication costs in Denmark provides an itemization of those costs on a national level, categorizing activities such as writing and peer review into billions of Danish kroner. The study found that open access publishing costs represented a relatively small percentage of the overall costs of scholarly communication: whereas "writing"<sup>i</sup> in the eight studied universities cost DKK 2.5 billion (4.7 million USD at the time of writing), article processing charges (APCs) cost DKK 230 million (43 million USD), and self-archiving cost only DKK 9 million (1.7 million USD). Ubiquity Press, an open access publisher, provides a similar view on a smaller scale, noting that of their costs, only 8% goes to "publishing, promotion, indexing & archiving."<sup>ii</sup> Kitchin, Collins, and Frost (2014) reference the PEER project, which found that initial setup of an institutional repository for green open access costs around \$60,000, with a broadly-ranging ongoing cost of \$2-53 per article. Finally, Burns, Lana and Budd (2013) provide a more comprehensive history of research into the cost of institutional repositories and green open access, including a breakdown of implementation and maintenance costs by size of repository and institution. Their findings suggest that the costs of institutional repositories vary most greatly by institution type and size, but are relatively unchanged based on solution (e.g., proprietary system versus open source software).

If the value of scholarship is in its dissemination, that value still varies based on the field, research culture, and location involved; the scholarly communication ecosystem is nothing if not complex. While the open access model has been developed to a greater extent within the sciences, and to suit the article publishing culture, it is still relatively nascent in the humanities

and social sciences. With this disparity in mind, OAPEN's work in the Netherlands gives insight into the costs of open access monograph publishing in the humanities—for example, they suggest that one book costs around €12,000 (\$16,000 in the year of publication) (Ferwerda, Snijder, and Adema 2013, 4). Eve's 2014 analysis of Ubiquity Press's costs provides the best per-article snapshot in the humanities, listing digital preservation and digital object identifier (DOI) costs as a lump charge of £20 (around \$34 at the time of writing) (4). In the sciences, PLOS (the Public Library of Science) charges fees per article that range from \$1,350 to \$2,900, depending on the field.<sup>iii</sup> As Eve suggests, there are possibilities of further cost savings in all of these estimates, but they could lead to a “race to the bottom” in more ways than one (4). Just as viable long-term economic models for open access are still largely in the development phase, the same is true of models for digital preservation. This slow progress is related to similar challenges in open access, where there are no direct incentives to undertake necessary actions, and the value of digital objects and their curation is not immediately realized (Grindley, 21). Additionally, costs are hard to identify, as many are carried within the institution's own cyberinfrastructure and must be projected into the long term as ongoing costs, with no way to know how those costs might evolve. That said, many projects have been and are currently being carried out. The European Union's 4C project forecasts that these projects will lead to a much more comprehensive understanding of digital preservation in the near term, as its vision states:

*In five years time (2020) it will be easier to design or procure more cost effective and efficient digital curation services because the costs, benefits and the business cases for doing so will be more widely understood across the curation lifecycle and by all relevant stakeholders.<sup>iv</sup>*

Notable current cost models include the DANS-ABC model, which was developed by the Royal Dutch Academy of Arts and Sciences (KNAW) Data Archiving and Networked Services (DANS) Institute, and which provides a project-based model for data storage in a trusted digital repository (Palaiologk et al. 2012).<sup>v</sup> In the US, the Total Cost of Preservation (TCP) project carried out by the California Digital Library also lays out the costs of digital preservation, and provides potential economic models based on those factors (Abrams et al. 2012). The UK-based LIFE project provides a deep analysis tool for digital preservation costs as well, in the form of a user-friendly spreadsheet for codifying costs based on activities in the curation lifecycle (Hole, et al. 2012). The 4C project has also developed a website for cost modelling and comparisons

entitled the Curation Costs Exchange.<sup>vi</sup> Kejser et al. (2014) provide a comprehensive analysis of other cost models.

While the cost models vary in granularity and focus, some shared assumptions are clear. Dedicated staff hours account for the majority of digital curation and preservation costs – between 50 and 70% (Kenny as cited in Houghton et al., 98; Beagrie, Chruszcz, and Lavoie as cited in Palaiologk et al., 196). Initial ingest carries large costs as well, and provides a solid area for potential cost offsetting and experimentation (Beagrie, Lavoie, and Wollard as cited in Hole et al., 4). The drop in storage costs, known as Kryder’s Law, is leveling out and can no longer be counted on as a future cost-saving measure (Rosenthal et al, 2), especially with storage diversity being seen as more and more intrinsic to the security of digital objects (Hole et al, 4). As cost modeling continues to develop and the infrastructure of projects becomes more advanced, these shared assumptions will surely continue to converge, just as the 4C project anticipates. This convergence will allow more institutions to integrate active digital preservation into their missions and share resources and costs in collaborative efforts.

### **Digital Preservation of Open Access Materials**

While a full treatment of the relevant technical aspects of digital preservation could span many book chapters, some of the most basic elements should be addressed briefly here to preface the discussion. To take a broad view, digital preservation encompasses all aspects of cyberinfrastructure: organizational (economics and staff), technical (systems and software), and administrative (policies). The Blue Ribbon Task Force for Sustainable Digital Preservation and Access (2010) found that sustainability revolved around six key items: ongoing allocation of resources, appropriate governance, incentives to act, understanding of value by decision makers, proper selection, and timely interventions for digital preservation (74).

The organizational element of digital preservation involves the financing of activities: how funds are allocated and whether preservation is taken as an ongoing cost or as a series of smaller interventions. The organizational aspect also largely involves staff, both those dedicated to the curation of digital objects and those who are tangentially related. Technical means of supporting long-term preservation include diverse storage measures involving different formats and locations, the use of open standards and systems, and effective preservation metadata. While there is no final answer regarding how many copies of an item should be kept, it is good practice

to have one or more mirrored copies. Further, geographically diverse storage is important in case of a catastrophic disaster; many institutions take part in collaborations like LOCKSS<sup>vii</sup> to ensure proper storage conditions. Also, in the case of extended digital preservation, it is extremely important to utilize open standards and systems in order to prevent content lockdowns or technological obsolescence due to the use of proprietary formats or software. This issue has largely been taken up in the library community in terms of technical systems and repositories, and some of the most popular efforts to deal with specific concerns relating to the preservation of open access materials will be addressed below. Finally, in order to make the format, systems, and software comprehensible, preservation metadata is needed. This includes a variety of information such as access and use rights, fixity, technical specifications, and versioning/provenance, all of which will be addressed later in the discussion of authenticity. Digital preservation must also be supported by administrative policies governing curation activities, just as collection development policies govern physical collections. Such policies should outline the scope and aim of a preservation program as well as the supported content types and formats, allowing staff to undertake any activities—whether migration, emulation, or deselection—necessary to provide stewardship of the content. Further, a policy outlining ongoing stewardship in case of departmental or institutional closure is important to ensure the extended viability of digital collections. While libraries regularly address the need for policies to support their mission and aims, only half of the respondents to Li and Banach’s survey of ARL institutions had preservation policies in place for their institutional repositories. OpenDOAR, a directory of international open access repositories, lists only 7.5% of its repositories (192 out of 2,561) as having a defined preservation policy.<sup>viii</sup> A lack of strategic development can be catastrophic for digital preservation efforts at worst, and costly at best.

It is important here to mention that even with the best of intentions, digital preservation plans can sometimes result in failure. Just as we know nothing of the future users of the content we are working to make accessible, we also cannot know the intentions and abilities of future stakeholders. At best, we can strategize and make preservation as easy as possible for the next generation, and hope that they will do the same. In the meantime, it is important to understand the breakdown of roles and responsibilities, as well as the various mature models of open access, in order to discover how they might affect long-term sustainability.

## **Stakeholder Roles**

Because the open access ecosystem includes a broad variety of stakeholders, the concept of ownership and responsibility in preservation is quite diffuse. The challenge here is twofold: determining, first, who is responsible for maintenance and preservation, and second, the impact of content ownership on preservation. These questions are not readily answered, as the stakeholder who will gain benefit from the long-term preservation of a particular material is most likely a separate entity from the actual “owner” of the content who “controls [its] long-term disposition” (Lavoie 2004). Determining the roles of stakeholders is also a complicated endeavor, because there is no precedent in the traditional scholarly record. There is great possibility for more interaction between stakeholders within the system, but it remains to be seen whether there will be a cohesive acceptance of responsibility, or if only a few institutions will shoulder a majority of the burden of preserving open access materials.

In many of the discussions below of stakeholder actions to preserve open access content, a trend appears that goes beyond the scope of any one group. It is a dependence on what are commonly referred to as “trusted third parties” or “trusted third-party archives,” which include “webscale”<sup>ix</sup> repositories like PubMed Central, Europeana, and the Internet Archive; community collaborations like LOCKSS, Academic Preservation (AP)Trust, and the Digital Preservation Network (DPN); and trusted vendor solutions like Portico. While webscale repositories are generally fully funded either by national or international entities or via grants, community initiatives are built on a member model and often utilize resources from members in order to build their services. Though webscale repositories and community-based initiatives approach digital preservation from different angles and with different resources – where webscale repositories have generous plans for accepting content and large buy-in from national and international entities, community-based projects are able to create an infrastructure of knowledge sharing that allows members to develop their own capacities alongside it – the two approaches are generally seen as having both the focus and the means for long-term preservation of items. All of these solutions provide digital preservation in very different ways and at varying cost levels. In the case of open access content, the choice of a third-party archive is relatively limited; Corrado and Moulaison posit that many open access publishers and journals are left out of paid initiatives due specifically to the cost involved (5). This exclusion effectively reduces the available options to the large repositories like PubMed Central and the Internet Archive, which, while extremely successful, can possibly place limits on preservation activities (e.g., publishers

might be limited in format or field). Beyond these enormous repositories, there is growing support for open access as the national and international infrastructures for digital preservation are built out. One example is the Public Knowledge Project's Open Journal Systems (OJS), which is currently developing a Private LOCKSS Network in order to preserve all of the journals published on the platform.<sup>x</sup> As a highly implemented system, OJS's move to create a better preservation infrastructure signals the fact that more open access materials will be stewarded into the future with ease. Zenodo,<sup>xi</sup> a repository service integrated with the European Union's OpenAire infrastructure, provides further infrastructure for those projects and publications that may not have a place in a webscale repository or that have financing to support community or third-party preservation. Described as part "rescue repository for content that does not fit elsewhere" and part "publication vehicle for new scholarly work products" (Lavoie and Malpas 2015, 21), and having no restrictions as to format, Zenodo provides services to the "long tail" of the scholarly record.

To return to individual stakeholders after having outlined a broader trend, in the case of libraries as content "owners" there is a relatively high rate of acceptance of preservation responsibilities but a varied practical response. As mentioned, Li and Banach's survey uncovered the fact that only half of respondents had preservation policies in place, even though nearly all saw preservation as a function of the repository. This lack of a tactical approach is potentially dangerous, as institutional repositories are in most cases the final stop for open access content. Because libraries are avid participants in collaborations and national repositories, however, there are potential failsafes in place. Academic libraries are more regularly joining community collaborations like AP Trust and DPN, the latter of which currently has over fifty members. While such collaborations provide some sense of increased preservation focus within library digital collections, the burgeoning nature of both organizations, along with the initial cost involved (which, for DPN, covers twenty years of storage and could thus be problematic for annual library budgets), suggests that many institutions might be selective in the content they are putting in; this result remains to be seen, and will provide an area for interesting analysis once the projects are further underway and content has been ingested at scale.

For another stakeholder in the scholarly communication ecosystem -- university presses -- innovation has been a key concept for the past ten years, as profits have dipped and competition from commercial publishers has increased. Because these presses serve the university's mission



alongside the library's, they offer a promising opportunity for collaborative efforts to preserve backlogs of content, especially as print-on-demand services are being viewed as an area of great economic potential (Greco 2010, 11). One example is Michigan Publishing at the University of Michigan, where the concept of sustainability creates a holistic approach to dissemination, preservation, and access; they are focused on providing open access content wherever possible, and maintain a partnership with HathiTrust to preserve and provide access to press backlogs.<sup>xii</sup> A similar example from the University of Minnesota Press, however, illustrates the challenge of digital content management at a university press: their open access materials are provided through Google Books (Hayes and Holley 2014, 75), which suggests that it may be difficult to undertake digital preservation efficiently outside of a commercial endeavor.

Non-profit publishers and organizations seem to take a strong stance in favor of digital content preservation. For example, JSTOR, a not-for-profit organization that collaborates broadly with academic institutions, includes in their mission a detailed statement outlining digital preservation responsibility and stewardship. While explaining their own practices and responsibilities, JSTOR also mentions the potential of transfer to a third-party repository in the case of organizational failure.<sup>xiii</sup> In January 2015, Project MUSE, another not-for-profit organization that began as a collaboration of university presses, signed a contract with Portico to provide long-term preservation to over 600 journals.<sup>xiv</sup> These examples suggest that collaborative non-profit endeavors are often more focused on, and have greater knowledge to support, long-term digital preservation.

Commercial publishers who have moved into open access, such as Elsevier and Wiley, seem to vary in their response to digital preservation. Elsevier's main website outlines their digital preservation practices, which include CLOCKSS (Closed LOCKSS) and Portico memberships,<sup>xv</sup> as well as agreements in place with the Dutch National Library (KB), Library of Congress, and the British Library<sup>xvi</sup> to make content available in perpetuity. One interesting project of note is JoVE,<sup>xvii</sup> the Journal of Visualized Experiments, a formerly gold open access journal of video "articles" that turned into a subscription-based publication with delayed open access. JoVE sends all of its published content to PubMed Central to be released open access after a two-year embargo period. While the journal does not have any specific preservation policies in place, they view their relationship with PubMed Central as a means of assuring long-term access to their content (Jachtorowicz, A. Private correspondence, June 12, 2015). Here, open access and digital

preservation act to benefit each other: where JoVE's aim is to provide their content in open access format, PubMed Central's services result in both preservation and access. It is not clear whether this sort of incentivizing would work with other publishers, but this is another area for future research.

On the whole, stakeholders in the open access ecosystem are approaching digital preservation in very different ways, based on individual capacities. The type of publisher is not necessarily indicative of responsibility in preservation activities; rather, it seems that infrastructure, economic resources, and institutional knowledge of the digital preservation landscape are much more indicative of better practices. Partnerships are the way forward for any type of open access stakeholder, as they allow individual institutions to fill gaps and better spread out the responsibilities and costs of digital preservation.

### **Open Access Models and Digital Preservation**

Where collaborative efforts and institutional agreements are filling gaps in the open access stakeholder landscape, there is a distinct difference in the types of problems posed by the open access models employed. As Suber (2012) suggests, two of the main varieties of open access, green and gold, are differently (and complementarily) suited to fulfill the main functions of scholarly journals. Where green open access is more suited to "registration (time stamps)" and "archiving (preservation)," gold open access is better equipped to manage "certification (peer review)" (62). These characteristics map to some of the challenges faced in digital preservation of content under these two open access models: where green faces challenges of curation and authenticity (both arguably tied to certification), gold faces hurdles of financial incentives and lack of preservation infrastructure.

### **Green Open Access**

While many argue that self-archiving is the most cost-effective and sustainable model of open access (Houghton and Swan 2013, as cited in Bjork et al. 2013, 238), it nevertheless poses a few key challenges to the long-term preservation of digital scholarship, due largely to the decentralization inherent in self-archiving; where traditional publishing aggregates content with libraries and publishers, self-archiving can be undertaken through a variety of different methods. Many researchers archive content with their university's institutional repository, but others might place copies on personal websites or research pages. Further, where before the published version of an article or book was the only available copy, green open access allows for the proliferation

of different versions of one piece of scholarship. An author may archive a pre-print with their institutional repository, publish with a scholarly publisher, and then post a final copy to their Academia.edu profile, creating a complicated web of versions and locations that make thoughtful curation much more difficult.

As green open access supports the diffusion of digital scholarship on a much broader scale than traditional publishing, curation is also problematic. Digital curation is a process of active selection and stewardship and, as mentioned above, it connects the concepts of access and preservation in many ways. While it involves all areas of management of digital content, one step of the process worth noting is that of deselection, either through deletion or proactive collection development. So far, a lack of policies in digital collection development has led in many cases to one of two scenarios: either an institution has collected everything and faces an overabundance of materials with varying quality, formats and connected metadata; or else an institution has been overly strict regarding requirements such as formats collected (e.g. only PDF or PDF/A) and now holds a non-representative amount of content. “Preservation into perpetuity” is a diffuse concept, and many libraries charge their institutional repositories and digital preservation programs with accomplishing it while maintaining little understanding of what it actually means to them. With no policies to guide the development of digital scholarship collections, more liberal institutions or those with more resources might end up in the first scenario, while conservative or underfunded entities will likely have the second outcome. Another area of note is again the issue of decentralization. Curation requires knowledge of where content is and how it is being managed, but such knowledge is hardly possible with the diffusion of content on various platforms. In many instances, the onus for long-term digital preservation thus shifts to the content owner, rather than a preservation-focused manager. Unless a form of digital scholarship situates itself directly under an active collecting area for a library or repository, it is highly unlikely that it will ever be preserved by a third party without the content creator seeking out that preservation. Decentralization also complicates the concept of authenticity, in terms of both authority and preservability. As the 2013 report from the Open Access Monographs in the Humanities and Social Sciences conference outlines, those working within academia are still skeptical of the “reliability” of open access materials (5). Largely due to misconceptions about green open access, their skepticism is still of note with regard to digital preservation, since it relates to an area of shared interest that is ripe with new possible solutions,

most of which are technical. While preservation is seen as part of the institutional repository's purpose, the most widely used repository systems have not yet been developed with the functionality needed to make good preservation practices systematic and ease the human burden. DSPACE, for example, has not been viewed favorably by preservationists because of its out-of-the-box lack of preservation support, including specifically its issues with "PREMIS implementation, versioning, and format migration" (Mannheimer, et al. 2014). Fedora 4, the most recent iteration of another widely used open source repository system, addresses the issue of versioning and authenticity of digital content by allowing a manager to create and save "snapshots," as well as view and revert to historical versions. At the creation of the repository, Fedora 4 offers a variety of choices for version control, so that provenance can easily be tracked according to an institution's needs and finances.<sup>xviii</sup> Within other open source publishing venues, version tracking can also be problematic. Code4Lib Journal, for example, publishes on a free WordPress instance, and while WordPress allows authorized users to view update logs of individual pages and articles, it does not allow visitors to see updates.<sup>xix</sup> This limitation creates a challenge in providing readers with information regarding any updates to articles; while editors generally make a note of changes in the article's text, technical support allowing access to versioning information would create a much stronger sense of transparency and thus reliability.

### **Gold Open Access**

The gold open access model poses complications more similar to those of traditional publishers. Gold open access formats map closely to those coming out of traditional publishers, largely finished journals and books, so curation and decentralization are not challenges as they are with green open access. Rather—since the incentive to preserve content digitally is not directly felt, and the mechanisms in place to preserve green open access content (i.e. repository frameworks) are not as readily available—the main issues are financial sustainability and preservation infrastructure.

Since knowledge of distant costs is vague at best, creating a sustainable business model for the digital preservation of gold content is immensely difficult. This is especially true for journals and publishers that function under the article processing charge (APC) model, where a one-time fee is calculated to replace the revenue that might have been generated through purchase or subscription. As the incentive is arguably quite low to include digital preservation in the total costs, it becomes an even more difficult burden to shoulder without proper cost projections.

Whereas green open access relies upon repositories that, while often geared towards access, can support preservation, gold open access materials can be published in a variety of formats and on different platforms that are much less likely to provide any sort of preservation functionality. Because publishers often use their own in-house databases to store and manage content, the onus is again on the individual content creator to either build out their own technical system to include the needed support for preservation activities, or partner with an outside service. Community-focused projects such as the previously mentioned Open Journal Systems are more likely to move into the realm of digital preservation, since development is pushed by a variety of stakeholders and the incentives to act are direct; preservation-focused institutions aiming to create and manage open access journals are more likely to use a platform if it has the needed functionality.

### **Conclusion**

As the broader infrastructure to support both digital preservation and open access matures, with solutions being shared amongst constituents and partnerships helping to fill knowledge gaps, the role of stakeholder and model will become less relevant. In its current state, the ecosystem is becoming more geared towards ongoing preservation, but there are visible disparities. Libraries taking charge of content must enact strong preservation policies and robust preservation-oriented technical systems. Publishers, both dedicatedly open access and not, must accept some level of responsibility for long-term preservation and provide stewardship plans to include either partnerships with libraries or input into third-party repositories. Collaborations between stakeholders will be key as the landscape matures and as individual institutions are able to pool resources and knowledge. All stakeholders should aim to participate in collaborations and work with third party repositories in order to ensure both preservation and access to the scholarly record.

While all of this discussion suggests that, from a preservation perspective, choosing open access publishing over traditional models has relatively little impact, the area of greatest concern (which is also an issue with traditional materials) is the ‘long tail’ of digital scholarship--those objects, publications, and collections that are not included in large content pushes to third-party archives, national libraries, or universities. These materials are notable because they are difficult to corral into repositories, which makes the scale of the challenge quite hard to determine. Just as with traditional publishing, lesser-known and -utilized scholarly objects still present a great

opportunity to researchers in the distant future. We must work to build greater inclusivity into digital preservation practices, in order to support the increased access that OA affords.

While we have seen vast developments in the infrastructure for digital preservation over the past decade, new challenges continue to emerge. We are now moving towards tackling issues of scale and complexity, such as preservation and access to large audio-visual collections. With the more straightforward challenges having been worked through, we are reaching some of the core issues that face our community when the scholarly record becomes increasingly digital. These issues map closely to those faced by the open access community, since content that presents preservation challenges often also tests the limitations of new and innovative publishing models. These challenges cannot be addressed by practitioners or institutions acting alone, and will not be solved quickly. Again, collaborative efforts are key in the further evolution of the open scholarly publishing landscape.

In his 1990 article “To Remember and Forget: Archives, Memory, and Culture,” Kenneth Foote compared the stewarding of crucial information regarding nuclear waste with the building or effacing of memorials to commemorate tragedies. Foote used this example to illustrate the fact that memory institutions, while ostensibly apolitical and objective, are often involved in a sort of identity-building that shapes and reinterprets the past. In the case of the scholarly record, this is an important point with which to conclude. Open access and effective digital preservation present new opportunities to build the evolving scholarly record. In the academic library, as stewards of that record, we must be mindful of the opportunities and responsibilities this role carries. We have a great opportunity to contribute to the evolving scholarly record, and to steward it into the very distant future.

## References

- Abrams, Stephen et al. Total Cost of Preservation (TCP): Cost Modeling for Sustainable Services. Los Angeles: University of California Curation Center, 2012. Retrieved from <https://wiki.ucop.edu/download/attachments/163610649/TCP-total-cost-of-preservation.pdf?version=5&modificationDate=1336402730000>
- Bjork et al. Anatomy of Green Open Access. *Journal of the Association for Information Science and Technology*, 65:2, 237-250, 2014. DOI: 10.1002/asi.22963
- Blue Ribbon Task Force for Sustainable Digital Preservation and Access. Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information. 2010. Retrieved from [http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf)
- Burns, C. Sean, Lana, Amy, Budd, John M. Institutional Repositories: Exploration of Costs and Value. D-LIB Magazine, 19, 1/2, 2013. Retrieved from <http://www.dlib.org/dlib/january13/burns/01burns.html>
- Corrado, Edward and Moulaison, Heather Lea. Digital Preservation for Libraries, Archives, and Museums. Lanham, MD: Rowman & Littlefield, 2014.
- Eve, Martin Paul. All that Glitters: Investigating Collective Funding Mechanisms for Gold Open Access in Humanities Disciplines. *Journal of Librarianship and Scholarly Communication*, 2:3, 2014. Retrieved from [http://eprints.lincoln.ac.uk/14337/1/\\_\\_ddat01\\_staffhome\\_bjones\\_Downloads\\_All%20That%20Glitters\\_%20Investigating%20Collective%20Funding%20Mechanisms%20fo.pdf](http://eprints.lincoln.ac.uk/14337/1/__ddat01_staffhome_bjones_Downloads_All%20That%20Glitters_%20Investigating%20Collective%20Funding%20Mechanisms%20fo.pdf)
- Ferwerda, Eelco, Snijder, Ronald and Adema, Janneke. OAPEN-NL: A Project Exploring Open Access Monograph Publishing in the Netherlands. The Hague: OAPEN Foundation, 2013. Retrieved from [http://www.researchgate.net/profile/Ronald\\_Snijder/publication/273450006\\_OAPEN-NL\\_-\\_A\\_project\\_exploring\\_Open\\_Access\\_monograph\\_publishing\\_in\\_the\\_Netherlands\\_Final\\_Report/links/5501fcd60cf2d60c0e629779.pdf](http://www.researchgate.net/profile/Ronald_Snijder/publication/273450006_OAPEN-NL_-_A_project_exploring_Open_Access_monograph_publishing_in_the_Netherlands_Final_Report/links/5501fcd60cf2d60c0e629779.pdf)
- Grindley, Neal. The Digital Curation Sustainability Model (DCSM). London: JISC, 2015. Retrieved from <http://www.4cproject.eu/dcsn>
- Hayes, Clayton and Holley, Robert. The University Press: Trends, Initiatives and Collaborations Over the Past Several Years. *Collection Building*, 33, 3, 2014. DOI 10.1108/CB-03-2014-0016
- Hess, Charlotte and Ostrom, Eleanor (Eds.). Understanding knowledge as a commons: from theory to practice. Cambridge, MA: MIT Press, 2007.
- Hole, Brian, et al. LIFE3: A predictive costing tool for digital collections. In: 7th International Conference on Preservation of Digital Objects (iPRES2010), 19 - 24 Sep 2010, Vienna, Austria.
- Houghton, John. Costs and Benefits of Alternative Publishing Models: Denmark. Melbourne: Victoria University, 2009. Retrieved from <http://www.knowledge->

[exchange.info/Files/Filer/downloads/DK\\_Costs\\_and\\_benefits\\_of\\_alternative\\_publishing\\_models.pdf](http://exchange.info/Files/Filer/downloads/DK_Costs_and_benefits_of_alternative_publishing_models.pdf)

Houghton, John et al. Economic Implications of Alternative Scholarly Publishing Models: Exploring Costs and Benefits. London: JISC, 2009.

Kejser, et al. D3.1 – Evaluation of Cost Models and Needs & Gaps Analysis. Copenhagen: 4C Project, 2014. Retrieved from <http://www.curationexchange.org/docman/4c-deliverables/13-d3-1-evaluation-of-cost-models-a-needs-and-gap-analysis/file>

Kitchin, Rob, Collins, Sandra, and Frost, Dermot. Funding models for Open Access digital data repositories. Online Information Review, 39, 5, 2015.

Lavoie, Brian. Of Mice and Memory: Economically Sustainable Preservation for the Twenty-First Century. In Access in the Future Tense. Washington, DC: CLIR, 2004.

Lavoie, Brian and Malpas, Constance. Stewardship of the Evolving Scholarly Record: From the Invisible Hand to Conscious Coordination. Dublin, OH: OCLC Research, 2015. <http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-esr-stewardship-2015.pdf>

Li, Yuan and Banach, Meghan. Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries. D-LIB Magazine, 17:5/6, 2011. <http://www.dlib.org/dlib/may11/yuanli/05yuanli.html>

Mannheimer, S et al. A Balancing Act: The Ideal and the Realistic in Developing Dryad's Preservation Policy. First Monday, 19:8, 2014. <http://www.firstmonday.dk/ojs/index.php/fm/article/view/5415/4105>

Open Access monographs in the humanities and social sciences [OAMHSS] conference. London: JISC, 2013. Retrieved from: <https://www.jisc-collections.ac.uk/Documents/Reports/OA%20Monographs%20in%20HSS%20Conference%20Report%20Oct%202013.pdf>

Palaiologk, Anna S. et al. An Activity-Based Costing Model for Long-Term Preservation and Dissemination of Digital Research Data: The Case of DANS. International Journal of Digital Libraries, 12: 195-214, 2012. DOI 10.1007/s00799-012-0092-1

Rosenthal, David S. H et al. The Economics of Long-Term Storage. In The Memory of the World in the Digital Age: Digitization and Preservation. Vancouver, BC: UNESCO, 2012. Retrieved from <http://www.lockss.org/locksswp/wp-content/uploads/2012/09/unesco2012.pdf>

Snijder, Ronald. Evaluating the Impact of the FWF-E-Book-Library Collection in the OAPEN Library: An Analysis of the 2014 Download Data. D-LIB Magazine, 21:7/8, 2015. <http://dlib.org/dlib/july15/snijder/07snijder.html>



Suber, Peter (2012). Open Access. Cambridge, MA: MIT Press.  
[https://mitpress.mit.edu/sites/default/files/9780262517638\\_Open\\_Access\\_PDF\\_Version.pdf](https://mitpress.mit.edu/sites/default/files/9780262517638_Open_Access_PDF_Version.pdf)

Tatum, Clifford and Jankowski, Nicholas. Beyond Open Access: A Framework for Openness in Scholarly Communication. In Wouters, Paul, et al. (Eds.) Virtual Knowledge: Experimenting in the Humanities and Social Sciences. Cambridge, MA: MIT Press, 2013.

---

<sup>i</sup> Writing in the case of this study involved direct time spent writing. It did not include pre-writing activities like research or post-writing activities in the publication process.

<sup>ii</sup> <http://www.ubiquitypress.com/site/publish/>

<sup>iii</sup> <https://www.plos.org/publications/publication-fees/>

<sup>iv</sup> <http://4cproject.eu/roadmap>

<sup>v</sup> DANS is also the home of the Data Seal of Approval, a repository certification service, and has helped to pioneer the concept of ‘trusted’ in developing preservation services.

<sup>vi</sup> <http://www.curationexchange.org/understand-your-costs>

<sup>vii</sup> <http://www.lockss.org/>

<sup>viii</sup> <http://opendoar.org/onechart.php?cID=&ctID=&rtID=&clID=&lID=&potID=5&rSoftWareName=&search=&groupby=pog.pogHeading&orderby=pog.pogID&charttype=pie&width=600&height=300&caption=Recorded%20Preservation%20Policies%20-%20Worldwide>

<sup>ix</sup> ‘Webscale’ in this case refers to repositories involving a national or international faction of content contributors and users, and which are built to scale with the continuous input of content. For a deeper definition, see Lorcan Dempsey’s description of the term:  
<http://orweblog.oclc.org/archives/001238.html>.

<sup>x</sup> <https://pkp.sfu.ca/pkp-lockss/>

<sup>xi</sup> <https://zenodo.org>

<sup>xii</sup> <http://deepblue.lib.umich.edu/bitstream/handle/2027.42/110792/ACRL%202015%20mkahn%20slides%20and%20remarks.pdf?sequence=1>

<sup>xiii</sup> <http://about.jstor.org/content/preserving-scholarship>

<sup>xiv</sup> <http://www.portico.org/digital-preservation/news-events/news/general-news/project-muse-to-preserve-e-journals-with-portico>

<sup>xv</sup> <https://www.elsevier.com/about/company-information/policies/digital-archive>

<sup>xvi</sup> <https://www.elsevier.com/physical-sciences/mathematics/perpetual-preservation-of-scholarly-content>

<sup>xvii</sup> <http://www.jove.com/>

<sup>xviii</sup> <https://wiki.duraspace.org/display/FEDORA40/Versioning>

<sup>xix</sup> <https://en.support.wordpress.com/pages/page-revisions/>